BAB II

LANDASAN TEORI

2.1 Robot Network (Botnet)

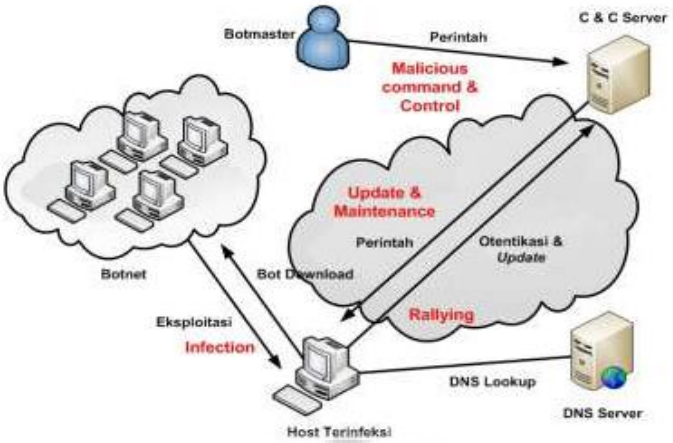
Botnet berasal dari kata “Bot” yang berasal dari kata “robot” yang artinya sebuah proses otomatis yang berinteraksi dengan layanan jaringan yang lain. Bot sendiri dapat beroprasi sebagai kegiatan positif atau negatif tergantung tujuannya.[3]

Botnet jika digunakan untuk hal positif adalah *indexing/spidering website* yaitu dapat mengetahui dan mengumpulkan harga-harga produk online sehingga dapat dibandingkan atau dapat menjalankan suatu fungsi sistem tertentu. *Botnet* yang dilakukan dalam tindakan negatif adalah menjalankan fungsi-fungsi tertentu dalam sebuah sistem komputer tanpa sepengetahuan pemiliknya bahkan dapat mencuri informasi penting dari komputer. Prilaku botnet dapat dianalogikan seperti koloni lebah, yaitu terdapat ratu sebagai pusat komando untuk memberi perintah kepada para lebah pekerja (komputer bot).[1]

Botnet merupakan salah satu ancaman yang serius terhadap keamanan diinternet, karena botnet dapat menyediakan *platform* yang didistribusikan untuk kegiatan ilegal serangan yang ada diinternet seperti *spaming, phishing, click fraud* dan *Distributed Denial of Service (DdoS).* Perbedaan *botnet* dengan *malware* lainnya adalah *botnet* dapat dikendalikan dari jarak jauh oleh *botmaster* dibawah infrastruktur *Command and Control (C&C) channel.* Perbedaan zona waktu, bahasa dan hukum membuat keberadaan botnet ini sulit untuk di lacak dan diketahui keberadaan hingga aktivitasnya.[6]

2.1.1 Siklus Hidup *Botnet*

Siklus hidup botnet terdapat 4 tahap menurut [6] yaitu : *infection, rallying, mallicious command and control, update and maintenance*



Gambar 2.1.1 Siklus Hidup Botnet

Sumber : [3]

Bot mampu mendistribusikan dirinya sendiri melalui jaringan internet dengan cara mencari celah terhadap komputer yang rentan dan tidak dilindungi untuk diinfeksi. *Malware* ini disebar oleh para *botmaster* dengan cara menyusupkan kesitus-situs legal yang berbahaya yang sudah dieksploitasi seperti instalasi software dari sumber yang tidak resmi, mengirimkan spam email agar korban mengklik link tertentu, hingga backdoor yang ditinggalkan oleh virus.

Setelah fase *infection* sukses, *host* yang terinfeksi akan mendownload *script code* dari sebuah *remote server* dan otomatis di install lalu merubah host tersebut menjadi *zombie.*

Proses *rallying* diawali dengan pencarian *C&C server* oleh host yang terinfeksi dengan menggunakan *DNS lookup.* Kebanyakan *botnet* menggunakan *Internet Chat Relay (IRC)* server xebagai *C&C server* untuk mengontrol botnet. Setelah mendapatkan alamat dari *C&C server* lalu *bot* melakukan *login* dan mengontentikasikan dirinya sebagai bagian dari botnet tertentu.

Pada *malicious command and control* botmaster bisa mengeluarkan perintah kepada bot untuk melakukan serangan spamming, click fraud, DdoS attack maupun menyebarkan *malware* agar dapat memperluas jaringan.

Botmaster juga dapat melakukan perintah update kepada *bot* untuk melakukan pembaruan melalui *download script code* agar dapat menambah

fungsionalitas *bot army*, menghindari pendeteksian serta memperbarui alamat dari *C&C server.*

2.1.2 Bentuk Serangan *Botnet*

Beberapa Serangan Botnet menurut [7] diantaranya :

a) *Spamming*, bot memiliki kemampuan membuka *sock proxy*, sebuah *proxy generik* untuk TCP/IP berbasis aplikasi jaringan pada mesin diincar setelah *sock proxy* diaktifkan, mesin ini dapat digunakan untuk kejahatan mengirim *spam* atau *email phising*

b) *Sniffing Traffic*, bot dapat menggunakan paket *sniffer* (penyadap paket yang melewati jaringan) untuk melihat data yang menarik di komputer sepertu *username* dan *password.*

c) *Distributed Denial of Service (DDoS) attack*, merupakan serangan jaringan yang mengakibatkan hilangnya layanan kepada *user*, hilangnya konektifitas jaringan dengan mengknsumsi *bandwidth* dari korban jaringan atau *overloading* sumber daya sistem komputasi korban.

d) Penyebaran *Malware, botnet* juga daoat menyebarkan malware, hal ini karena bot menerapkan mekanisme untuk men*download* dan menjakankan file melalui *http* dan *ftp*, beberapa bot dapat berpura-pura berperan sebagai *server http* atau *ftp* sebagai malware.

e) *Installing Adversitement Add-ons & Browser Helper Objects*, membuat situs palsu dan beberapa iklan yang mendaftar *pay per clicks* di penyedia iklan, botmaster juga dapat pemasukan dengan bantuan daru sebuah botnet, proses ini dilakukan secara otomatis (klik penipuan) sehingga rubuan bot mengklik iklan tersebut.

2.2 Deteksi Botnet

Deteksi adalah usaha agar menemukan dan menentukan keberadaan, anggapan maupun kenyataan. Deteksi botnet dapat diartikan menjadi usaha untuk menemukan sebuah *botnet* dengan beberapa parameter tertentu didalam jaringan dengan menggunakan metode maupun teknik-teknik tertentu.[3]

2.2.1 Honeypots

Honeypot merupakan perangkap untuk umpan agar memikat *client* sebagai penyerang yang mencoba masuk secara paksa dalam suatu sistem. Honeypot digunakan untuk memantau dan mempelajari metode yang digunakan hacker untuk menembus suatu sistem. Informasi yang didaptkan akan digunakan sebagai tindakan preventif dimasa akan datang.[6]

2.2.2 Passive Monitoring on Network Traffic

Passive Monitoring on Network Traffic adalah teknik memonitor lalu lintas yang melewati suatu jaringan dan kemudian dianalisis untuk mengidentifikasi keberadaan dan memahami dari karakter botnet itu sendiri seperti *signature-based, anomaly-based, DNS-based,* dan juga *data mining-based.* [4]

a) *Signature-based,* mengidentifikasi paket yang melewati suatu jaringan lalu mencocokkannyadengan informasi yang telah diperoleh dari penelitian sebelumnya. Bagian yang diteliti berupa perilaku maupun ukuran dari paket. Signature sendiri merupakan perilaku ataupun kode yang secara unik dapat mengidentifikasi botnet tertentu. Namun *signature based* memiliki kelemahan yaitu tidak dapat diterapkan untuk mendeteksi *botnet* yang belum diketahui.

b) *Anomaly-based* merupakan pendeteksian botnet berdasarkan pada beberapa anomali lalu lintas jaringan seperti volume lalu lintas yang tinggi, lalu lintas di port yang tidak biasa dan adanya ancaman berbahaya dalam jaringan dari bot karena perilaku sistem yang tidak biasa. Dari kelakuan sistem dan pola pemakaianlah dasarnya pendeteksian ini.

c) *DNS based* merupakan informasi dari DNS yang dihasilkan oleh sebuah botnet. Teknik dari DNS-based ini mirip dengan anomaly-based. Biasanya bot melakukan koneksi dengan C&C server untuk mendapatkan perintah. Bot melakukan query DNS untuk mengakses C&C server. DNS based sangat memungkinkan untuk mengetahui botnet dari pemantauan dan juga deteksi lalu lintas dari DNS.

Passive DNS merupakan teknik replika yang secara pasif menangkap pesan DNS dan memingkinkan menyimpan data-data tersebut.

d) *Data Mining Based* model deteksi ini mengacu pada prosesnya yang otomatis untuk data yang jumlahnya besar. Proses data mining ini diterapkan untuk algoritma pada machine learning, pattern recognition machine dan juga yang lain-lain. Tujuan dari teknik ini adalah mendeteksi menyebaran dari botnet dan ditemukannya C&C server berdasarkan logfiles yang dikumpulkan.

2.3 Passive DNS

Passive DNS merupakan merupakan data yang terdiri dari beberapa jenis dengan fungsi yang berbeda yang dikumpulkan selama kurun waktu tertentu yang bertujuan untuk data yang dapat dianalisa demi kebutuhan tertentu. (Weimer, 2005)

Menurut [4] komponen dari passive dns untuk dapat mendeteksi sesuatu yang upnormal seperti botnet adalah *DNS-client, DNS-server, domain name, time stamp, client IP query type (RR), query name, answer, TTL*.

2.4 K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran sekumpulan data yang sudah terklasifikasikan sebelumya. K-Nearest Neighbor (kNN) termasuk dalam supervised learning, dimana hasil query instance yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam kNN. [5]

Keunggulan metode *K Nearest Neighbor* adalah metode ini dapat bekerja secara nonlinear dan nonparametric sehingga tidak perlu melakukan distribusi instance dalam dataset.

K Nearest Neighbors melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran, setiap data pembelajaran direpresentasikan menjadi titik-titik pada ruang dimensi banyak.

Data baru yang diklasifikasi selanjutnya diproyeksikan pada ruang dimensi banyak yang telah memuat titik-titik data pembelajaran. Proses klasifasikasi dilakukan dengan mencari titik terdekat dari data baru atau nearest neighbor. Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan

formula jarak Euclidean. Menurut [8] langkah-langkah untuk menghitng metode kNN adalah sebagai berikut :

a. Menentukan parameter *k*

b. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan

c. Mengurutkan jarak yang terbentuk

d. Menentukan jarak terdekat sampai urutan *k*

e. Memasangkan kelas yang bersesuaian

f. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2}$$ [8]

Keterangan:

x1 = Sampel data

x2 = Data uji atau data testing

i = Variabel data

d = Jarak

p = Dimensi data

Teknik pencarian tetangga terdekat disesuaikan dengan dimensi data, proyeksi, dan kemudahan implementasi oleh pengguna. Untuk menggunakan algoritma k nearest neighbors, perlu ditentukan banyaknya k tetangga terdekat yang digunakan untuk melakukan klasifikasi data baru. Banyaknya *k*, sebaiknya merupakan angka ganjil, misalnya k = 1, 3, 7, 9 dan seterusnya. Penentuan nilai *k* dipertimbangkan berdasarkan banyaknya data yang ada dan ukuran dimensi yang dibentuk oleh data. Semakin banyak data yang ada, angka k yang dipilih sebaiknya semakin rendah, namun semakin besar ukuran dimensi data, angka *k* yang dipilih sebaiknya semakin tinggi.

2.5 Bahasa Pemograman *Python*

Phyton adalah bahasa pemograman skrip yang berorientasi objek, phyton juga dapat digunakan untuk pengembangan berbagai perangkat lunak dan dapat berjalan

dibanyak platform sistem operasi. Phyton merupakan bahasa pemograman tidak ada batasan dalam penyalinan ataupun distribusinya. [9]

Beberapa fitur yang dimiliki Phyton diantaranya :

a. Memiliki *library* yang luas, dalam distribusi pun phyton telah disediakan modul-modul.
b. Memiliki tata bahasa yang mudah untuk dipelajari
c. Memiliki aturan layout kode untuk memudahkan pengecekan, pembacaan kembali maupun penulisan ulang kode.
d. Berorientasi pada obyek.

Phyton juga didistribusikan dengan beberapa lisensi yang berbeda dari beberapa versi. Namun pada prinsipnya Python dapat diperoleh dan dipergunakan secara bebas, bahkan untuk kepentingan komersial. lisensi Python tidak bertentangan baik menurut definisi Open Source maupun General Public License.

2.6 Kajian Tentang Penelitian Terdahulu

Pada penelitian yang dilakukan oleh "Didin Nizarul Fuadin" yang berjudul "Deteksi Botnet dengan Naïve Bayes dan Smote BFS" dikemukakan bahwa pengujian dengan dataset Botnet CTU-13, sebagai data uji model deteksi Botnet menggunakan Naïve Bayes dengan SMOTE dan metode BFS. Dengan data pengujian total 523.559 flow, yang berisi 517.887 flow trafik normal dan 5.672 flow trafik Botnet. Pada penelitian yang dilakukan oleh "Weikeng Robbin Chen" yang berjudul "Exploring a Service-Based Normal Behaviour Profiling System for Botnet Detection" mengemukakan bahwa penggunaan algoritma supervised learning machine dapat memberikan perkembangan pada hasil klasifikasi yang menggunakan sumber data dari CTU 13. Pada thesis Pedro Marques da Luz pada tahun 2013/2014 dengan judul "Botnet Detection Using Passive DNS" membandingkan pendeteksian berdasarkan real time dan juga tidak real time.

Perbedaan pembahasan yang dipaparkan oleh peneliti berupa pemantauannya di passive DNS yang berbeda dan metode yang berbeda dari penelitian yang dilakukan sebelumnya yaitu k-Nearest Neighbor dan melakukan pengujian di *passive DNS* yang didapatkan dari *dataset CTU-13* lalu diuji dengan K-Fold Cross Validation dan Compusion Metric.